Analysis of Multi-domain Complex Simulation Studies

James R. Gattiker, Earl Lawrence, David Higdon

Los Alamos National Laboratory
{gatt,earl,dhigdon}@lanl.gov

Abstract. Complex simulations are increasingly important in systems analysis and design. In some cases simulations can be exhaustively validated against experiment and taken to be implicitly accurate. However, in domains where only limited validation of the simulations can be performed, implications of simulation studies have historically been qualitative. Validation is notably difficult in cases where experiments are expensive or otherwise prohibitive, where experimental effects are difficult to measure, and where models are thought to have unaccounted systematic error. This paper describes an approach to integrate simulation experiments with empirical data that has been applied successfully in a number of domains. This methodology generates coherent estimates of confidence in model predictions, model parameters, and estimates, i.e. calibrations, for unobserved variables. Extensions are described to integrate the results of separate experiments into a single estimate for simulation parameters, which demonstrates a new approach to model-based data fusion.

1 Introduction

Computational simulation applications are increasingly used to explore a number of domains, including: climate, ocean, and weather modeling; atomic scale physics modeling; aerodynamic modeling; and cosmology applications. A significant challenge for using simulation studies is the quantitative analysis of simulation results, and the comparison and integration of the simulations with experimental data.

At Los Alamos National Laboratory, a challenge is to certify the safety and reliability of nuclear weapons where only indirect physical experiments can be performed[1]. Simulations model physical experimental results. Uncertainties arise from a variety of sources that include: uncertainty in the specification of initial conditions, uncertainty in the value of important physical constants (e.g., melting temperatures, equations of state, stress-strain relationships, shock propagation, and transient states), inadequate mathematical models, and numerical computation effects. Experimental observations constrain uncertainties within the simulator, and are used to validate simulation components and responses[10].

The methodology described here addresses three main goals in simulation analysis. First is the *quantification of uncertainty* in predictions. Most simulations systems lack the ability to directly assess the uncertainty in their results,

although it is clear from failure to match reality that both bias and uncertainty exist. The second goal is the *calibration of unknown parameters*. Simulations often have parameters that are either non-physical or are unmeasurable in experiments, and must be determined, or calibrated. The third goal addressed, discussed here for the first time, is the linking and *joint calibration* of variables common to separate experiments.

Additional issues constrain approaches to this problem. Experimental data in typical application areas is generally difficult to collect because of expense, difficulty in making physical measurements, or external constraints. Simulation studies are often computationally expensive, having usually been developed at the limits of feasible computation, and so there is limited access to alternative simulation parameter settings. Some exploration of alternatives is therefore possible, but putting the simulator directly into an iterated method can be prohibitive.

This paper describes a methodology that has been implemented and demonstrated to be effective in addressing these issues in real problems[2]. The approach is to model simulation response with an accurate emulated response. Parameters of this emulator as well as simulation parameters are simultaneously determined using a Bayesian parameter formulation and associated Markov chain Monte Carlo sampling. The emulator itself is a stochastic process model, modeling both simulation response and systematic bias.

1.1 The Model Evaluation Problem

This section follows an explanatory "toy" problem that captures many of the issues in analyzing (simulation) models in conjunction with experiments[3]. The task is to analyze models of gravitational attraction, through the combination of an analytical model and experiment. To study the nature of falling objects a test object is dropped from various heights to study their descent time. In addition to characterizing the measurements of limited experiments, we wish to extrapolate the behavior. In practice, the drop time is not a simple effect due to atmospheric effects. For explanatory purposes, experimental data is generated according to

$$\frac{d^2z}{dt^2} = -1 - 0.3\frac{dz}{dt} + \epsilon,$$

which includes the square law of gravitational attraction, plus a term for linear effects. If the simulation models only the gravitational attraction $\frac{d^2z}{dt^2} = -1$, the results do not explain the data well, and extrapolate even more poorly. This model would give a fitted response as shown in Fig. 1a. To analyze the simulation results, we need a model of the data that explicitly allows for systematic error in the simulation system. In this approach, we use a model η that models the simulation responses, and an additional model δ of the discrepancy between the simulation and the data, so that the model is comprehensive, i.e., $Y(z) = \eta(z) + \delta(z) + \epsilon$. The data can now be modeled accurately as a systematic discrepancy from the simulation. Incorporating uncertainty into the model

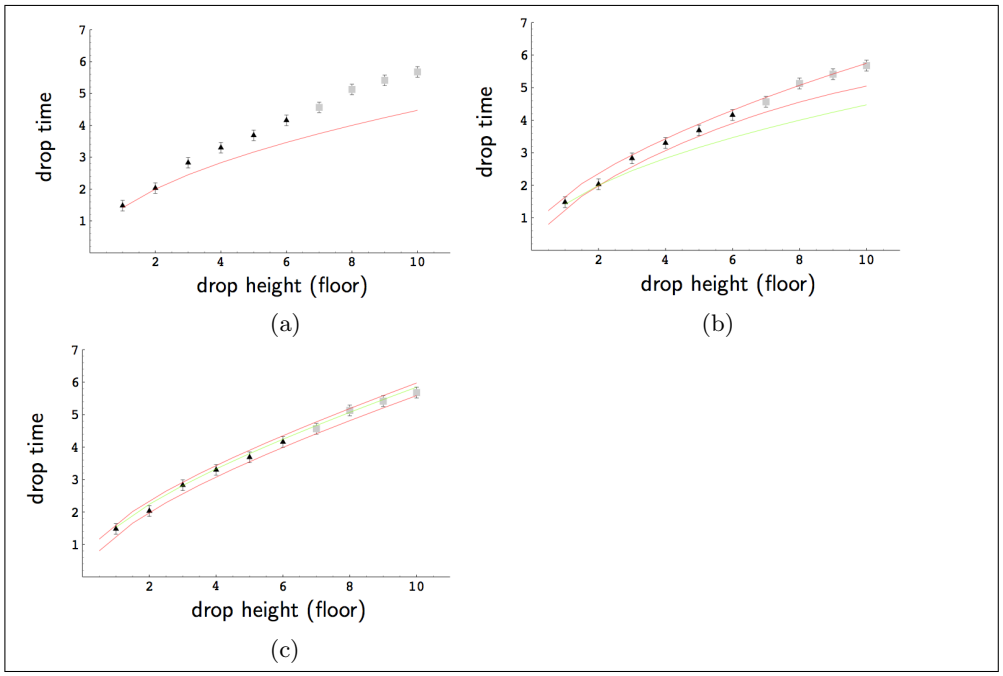


Fig. 1. Explanatory diagrams of experiment drop times. a) Results compared to an inadequate ideal model; b) model results and (bounded) discrepancy adjustment; c) revised model.

response, the results are shown in Fig. 1b. The model's η, or simulation-based prediction, remains distinctly low, while the discrepancy adjustment, with uncertainty incorporated, follows the data closely and responds better in extrapolation.

Continuing the analysis, the discrepancy term suggests the postulation of an improved model, for example:

$$\frac{d^2z}{dt^2} = -1 - \theta\frac{dz}{dt} + \epsilon,$$

where θ is an unknown to be calibrated. Determining this model, including uncertainty both in the model and the calibration of the parameter, the results are shown in Fig. 1c. In this case, our best simulation result with the calibrated parameter closely follows the data, and the complete model closely bounds the prediction. The enhanced model gives greater confidence in extrapolation as compared to the incorrect model relying on estimated discrepancy to fit the data.

To summarize, the problem starts with the modeling of simulation response, and the modeling of experiments as the simulation response plus some systematic bias. This discrepancy describes simulation model insufficiency (or other systematic bias). Parameters of these models are determined so that uncertainty

in predictions is implicitly given. Unknown parameters are calibrated to best reconcile the experimental data with the simulations, reporting on plausible values. Discrepancy estimates may be further used to examine problem foundations.

2 Model Formulation

This modeling approach was originally generated by Kennedy and O'Hagan [6], and has been discussed in application in [8, 9]. For space limitations, the model cannot be completely described here, but complete formulation of the single experiment configuration methodology is available [10, 11].

The core of the modeling effort is a Gaussian stochastic process model (GPM). These models offer an expressive method for modeling a response over a generic space [12]. The GPM relies on a specified covariance structure, called the *covariogram*, that describes the correlation between data points that scales with distance. In this case, nearby is measured as a scaled Euclidean distance in the simulation parameter space. The covariogram is then:

$$C(x_1, x_2) = \frac{1}{\lambda_z} exp^{-d(x_1, x_2, \beta)^2} + I\frac{1}{\lambda_s},$$

where $d = \sqrt{\sum \beta^i (x_1^i - x_2^i)^2}$. The λ parameters are precisions (inverse variances), with λ_z corresponding to the variability of the data, and λ_s corresponding to the variability of the residual. This correlation assumption constrains the response surface to a family functions. Predictions are made by computing a joint covariogram of the known and predicted datapoints, and producing the multivariate normal conditional distribution of the unknown locations. The predictions are then distributions rather than point estimates:

$$X_p \sim N(\mu_p, \Sigma_p).$$

This distribution can be used to produce realizations of possible values, and also allows the extraction of confidence bounds on the estimates.

η models the simulations, but a two-part model is used to also explicitly model the discrepancy δ between the observed data, y_{obs}, and the simulation response, y_{sim}, such that:

$$\begin{aligned} y_{sim} &= \eta(x, t) + \epsilon, \\ y_{obs} &= \eta(x, \theta) + \delta(x) + \epsilon. \end{aligned}$$

t are simulation parameters whose corresponding values are not known for the observed data. These unknown θ are determined (i.e., calibrated) in modeling, along with the β and λ parameters for for the η and δ models.

η and δ model parameters and θ values are generated with Markov chain Monte-Carlo sampling of a Bayesian posterior. The Bayesian approach gives the posterior density for the η model as

$$\pi(, \beta, \lambda | y_{sim}(x)) \propto L(y(x) | \eta(x, \beta, \lambda)) \times \pi(\beta) \times \pi(\lambda).$$

In words: the *posterior* distribution of the parameters given the simulation data is proportional to the likelihood of the data given the parameters times the prior probability of the parameters. The likelihood corresponds to a least squares measure of the model predictions compared to the given data, though this is not computed explicitly. This formulation has been extended to produce a single likelihood of the $\eta + \delta$ model fitting the observed and simulated data simultaneously [11].

The posterior of the parameters can be sampled from using Metropolis-Hastings MCMC sampler, resulting in samples from joint the posterior density of all parameters. It is possible to optimize directly on the likelihood function or the posterior for a point solution, but the resulting "optimal" solution has no associated confidence on the parameters. Computing the likelihood requires inversion of the $k \times k$ covariance matrix, where $k = n(p + q) + mq$, where n is the number of experimental data and m is the number of simulated data, and p is the dimensionality of the simulation response, and q is the dimensionality of the discrepancy response. This quickly becomes intractable if p grows large, so an important enhancement to the model is the use of linear basis dimension reduction in the output space, the effects of which can be compensated for in the modeling approach. Good results have been obtained with principle components reducing the p dimension, and kernel regression constraining and reducing q. This makes even problems with large data sizes, for example time series or images, computationally tractable [10].

3 Example Application: Flyerplate Experiments

In order to study the properties of materials under shock, plates of the material are subjected to a high-velocity impact. The velocity of the impacted plate is measured over time, revealing several material property driven regimes, as detailed in Fig. 2.

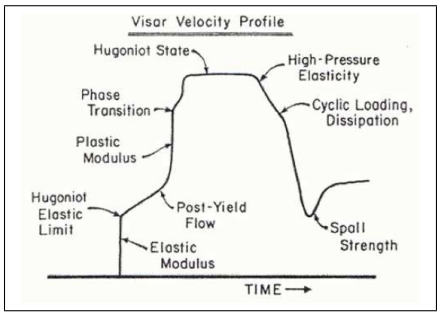


Fig. 2. Theoretical regions of flyerplate velocity measurements.

Results of flyerplate experiments are shown Fig. 3, which shows both the simulations from a 128 experiment design over 7 variables, and a trace of measured

data. The unknown θ parameters ranges have been defined by subject matter experts, and scaled to [0,1] for the purposes of the analysis. θ in this problem are parameters from the Preston-Tonks-Wallace stress-strain model[5]. Parameters include: θ_0 Initial strain hardening rate; κ material constant in thermal activation energy; $-\log(\gamma)$ material constant in thermal activation energy; y_0, maximum yield stress; y_∞, minimum yield stress; s_0, maximum saturation stress; s_∞, minimum saturation stress. The simulation data is modeled with the first five principal components. The discrepancy model is a kernel basis constraining relatively smooth variation over the parameter space, modeling an arbitrary general effect in the absence of a more specific model.

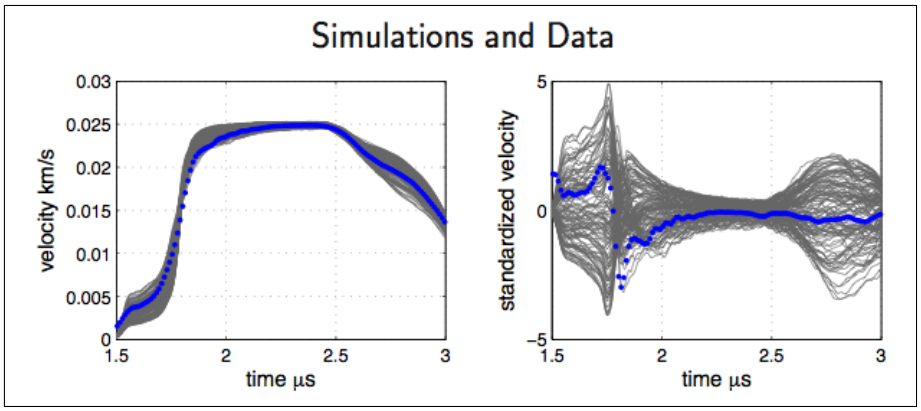


Fig. 3. Observed and simulated tantalum flyerplate velocity results, native data scale and standardized to mean 0 variance 1.

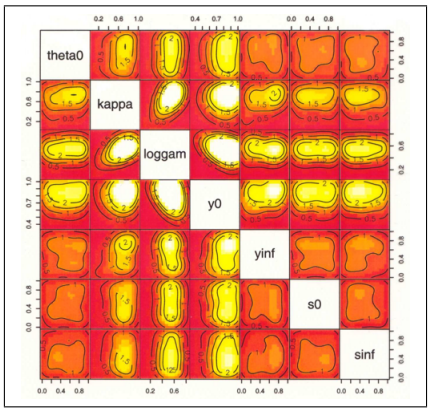


Fig. 4. Calibration results of unknown model parameters explored with simulations.

Figure 4 shows the results of the full joint parameter calibration as contours of the two-dimensional PDF projections, as sampled by the MCMC procedure. There are clear trends in some variables, which have been calibrated more tightly than their *a priori* ranges, whereas some do not have clear preferred values. An additional variable importance measure comes from the sampled spatial scaling parameters β, where lower spatial correlation corresponds to more active variables. This experiment verified that β_2, β_3, and β_4 are important, consistent with the θ calibration. Proper sensitivity analysis can be performed by analyzing model predictions, which can be produced cheaply from the emulator model. Figure 5 shows predictive results of the model in the scaled space. These predic-

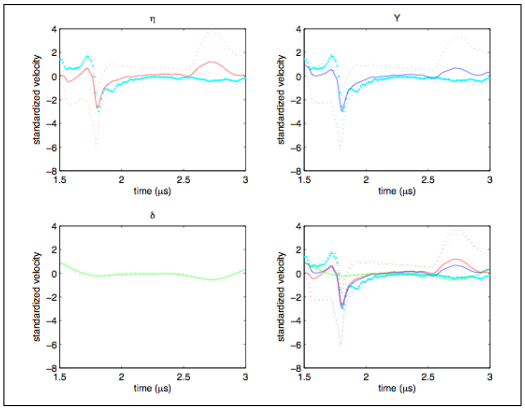


Fig. 5. Observed and calibrated predictions of the velocimetry results for tantalum flyerplate experiments (scaled). On the upper left is observed data in large dots, with the η model. Lower left shows the δ model contribution. The upper right is the observed data with the $Y = \eta + \delta$ result. The lower right plot repeats all results on one scale. Solid lines are mean response, and dotted lines show the associated 10%-90% confidence region of each quantity.

tions cover several model realizations (drawn from the Normal model distribution), predicted over many MCMC drawn parameter sets. The closest emulator simulation response is not able to capture the observed data, in particular in the edges where the simulations showed an initial higher-than-expected measurement, and late-time bump that is not observed in the data. The discrepancy adjusted response reduces this error. Of particular interest is that the simulator response alone not only failed to capture the observed data, but the confidence region is also not satisfactory. The $Y = \eta + \delta$ model's confidence regions are more appropriate.

4 Joint Calibration of Models

In complex applications, *separate effects test* are used to explore different aspects of a problem through surrogate experiments. Flyerplate results may be scientifically interesting in their domain, but the experiments are also intended to collect data on effects that are part of larger physical simulations. Some of these experiments will inform on the same parameters, and it is desired to perform a calibration that uses all of these results to simultaneously and consistently.

If two models have separate likelihoods $L_1(\theta_1, \theta_2|y)$ and $L_2(\theta_1, \theta_3|y)$, they can be considered a joint model as $L_J = L_1 \times L_2$. Using MCMC, the draws can be simplified, computing L_J for the draws related to θ_1, while for the parameters of independent models the likelihoods, L_1 and L_2 are computed independently for draws of θ_2 and θ_3, saving computation. This is a method to quantify the effects of variables in common between experiments.

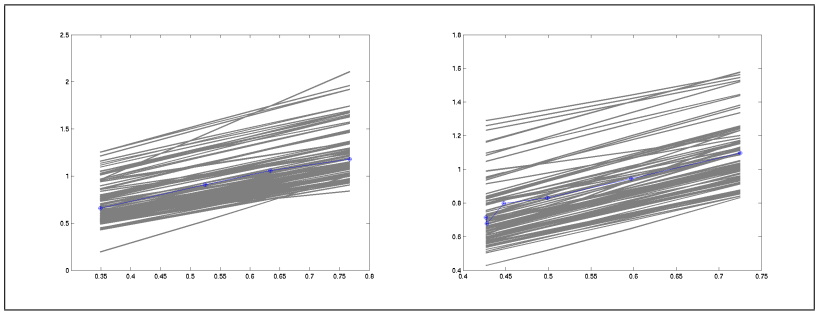
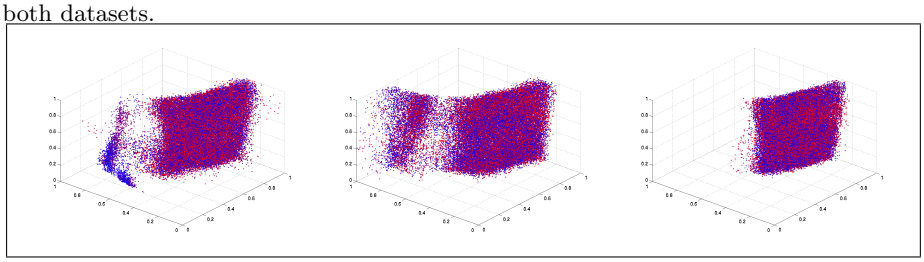


Fig. 6. Shock speed simulation models and measured data for hydrogen (left) and deuterium.

In the *shock speed modeling* problem, it is desired to model measured shock speed quantities of several materials. A single experiment in a single material consists of shock particle velocity u_p measured in response to shock of speed u_s. Several experiments characterize a material response curve, which is modeled by simulations, as shown in Fig. 6.

In the full application, there are many materials and several parameters, some of which are shared between models. We will limit the discussion to the characterization of hydrogen (H) and deuterium (D), which use the same parameters, referred to here as θ_1-θ_3. Figure 7 shows the calibration of parameters for each single model, as well as the joint model. The joint model shows a more compact and stable calibration, in a region that is expected by domain experts. These results show that this methodology can successfully capture data from different experiments and even different simulation systems to calibrate underlying parameters with greater fidelity than the single model calibrations are capable.

Fig. 7. Joint calibration of model parameters. The left plot shows the calibrated theta vectors for a hydrogen model alone, the middle plot shows the same parameters in a deuterium model, and the right plot shows the joint calibration of the parameters given both datasets.



5 Discussion

The approach described provides a method to quantify the implications of experimental data combined with simulation experiments. It is a tool to be used with domain experts from both the modeling and simulation domain, as well as the experimental data domain. Expert judgement is required in the generation of appropriate simulation studies, the construction of plausible discrepancy models, and the assessment of results. If domain knowledge is available to describe strong parameter priors, including θ parameter bounds and relationships, these may be incorporated, though by default weak priors that do not inappropriately constrain modeling and calibration can be used.

As is usual in complex models, attention to diagnostics is important. Because this modeling approach incorporates a number of trade-offs in the modeling space, it is possible that the model could enter an inappropriate domain, where it is fitting the data well, but the parameters are not plausible (e.g., counter to physics). Also, some expertise in MCMC is needed to identify the initial transient and to determine appropriate step sizes over the parameter ranges to ensure adequate mixing of the chain.

Without response smoothness, it is difficult to envision how to model and calibrate in this (or any) framework. Thus, a key issue in successful application is ensuring simulation response smoothness through the parameter space under analysis. If this is not an inherent quality of the data, variable transformations and feature construction studies are necessary.

In summary, this modeling approach:

- provides a method for integrating simulation experiments with empirical data, modeling systematic error in the simulation;
- calibrates unknown simulation parameters;
- provides well-founded uncertainty estimates in parameters and predictions; and
- allows separate experiment results to be fused into one result of parameter calibration.

When given two distinct datasets with a relationship in their underlying variables, it is generally not clear how to fuse the information in the datasets into a single answer. Information must be of the same type before it can be quantitatively reconciled, and this is usually solved by transforming the data directly into the same domain. The application methodology described here shows how different datasets may be linked by a generating model, in this case a simulation that can produce results in the various model domains. Through this approach, inverse problems from two distinct experimental domains can be combined, and a composite model realized.

References

1. *Los Alamos Science* special issue on Science-Based Prediction for Complex Systems, no.29, 2005.
2. J. Gattiker, "Using the Gaussian Process Model for Simulation Analysis Code", Los Alamos technical report LA-UR-05-5215, 2005.
3. Christie, Glimm, Grove, Higdon, Sharp, Schultz, "Error Analysis in Simulation of Complex Phenomena", *Los Alamos Science* special issue on Science-Based Prediction for Complex Systems, no.29, 2005, pp.6-25.
4. S. Chib, E. Greenberg, "Understanding the Metropolis-Hastings Algorithm", The American Statistician, Nov. 1995; 49, 4; p.327.
5. M Fugate, B Williams, D Higdon, K Hanson, J Gattiker, S Chen, C Unal, "Hierarchical Bayesian Analysis and the Preston-Tonks-Wallace Model", Los Alamos Technical Report, LA-UR-05-3935, 2005.
6. M Kennedy, A. O'Hagan, "Bayesian Calibration of Computer Models (with discussion)", Journal of the Royal Statistical Society B, 68:426-464.
7. D.Jones, M. Schonlau, W.Welch, "Efficient Global Optimization of Expensive Black-Box Functions", Journal of Global Optimization 13, pp. 455-492, 1998.
8. D. Higdon, M. Kennedy, J. Cavendish, J. Cafeo, and R. Ryne, "Combining field Observations and Simulations for Calibration and Prediction", SIAM Journal of Scientific Computing, 26:448-466.
9. C. Nakhleh, D. Higdon, C. Allen, V. Kumar, "Bayesian Reconstruction of Particle Beam Phase Space from Low Dimensional Data", Los Alamos technical report LA-UR-05-5897.
10. Dave Higdon, Jim Gattiker, Brian Williams, Maria Rightley, "Computer Model Calibration using High Dimensional Output", Los Alamos Technical report LA-UR-05-6410, submitted to the Journal of the American Statistical Association.
11. Brian Williams, Dave Higdon, James Gattiker, "Uncertainty Quantification for Combining Experimental Data and Computer Simulations", Los Alamos Technical Report LA-UR-05-7812.
12. D.R.Jones, M.Schonlau, W.Welch, "Efficient Global Optimization of Expensive Balck-Box Functions", Journal of Global Optimization 13, pp. 455-492, 1998.